DOCUMENT RESUME

ED 092 957 CS 201 388

AUTHOR Finder, Morris
TITLE Criterion Tests and Standardized Tests.
PUB DATE 74
NOTE 6p.; Paper presented at the Annual Meeting of the New York State English Council (24th, Binghamton, New York, May 2-4, 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS Achievement Tests; *Criterion Referenced Tests; *Educational Testing; Effective Teaching; *Standardized Tests; Test Results

ABSTRACT

Standardized tests are devised to sort students out, comparing them on a scale from high to low, not to discover what students know. Thus, results of standardized tests are misleading in representing the achievements of educational programs and in comparing one school or school system with another. Criterion tests, however, measure directly and specifically the intentions of teaching. These tests, based on the skills which are taught and written so that students may demonstrate the extent to which they have acquired the desired learning, provide the acceptable bases for improving both schools and teaching. (JM)

ED 092957

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Morris Finder
School of Education
State University of New York at Albany
Albany, New York 12222

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
Morris Finder
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

CRITERION TESTS AND STANDARDIZED TESTS[1]

To understand current controversies about testing, one must know

1. what normed or standardized tests are,
2. what criterion tests are, and
3. what the differences between these are.

One's knowledge of these distinctions is prerequisite to the intelligent conduct and judgment of schools and teaching.

Let us consider first the normed or standardized test. A brief historical note can help. In 1914 there were about 139,000 soldiers in the United States Army. With the onset of World War I, the army grew rapidly to about two million. Therefore the army was faced with an immense task of sorting. Who should go to officer candidate school? Who should go to cooks and bakers school? Who should be selected for this kind of job? And who for that? These decisions had to be made promptly and so psychologists were put to work on the problem. One result was the Army Alpha Test. When this test was administered to an unselected group, the army was told who was high, middling, and low on this particular instrument that measures some notion of mental alertness.

When the war had ended, some of the psychologists who developed these tests took jobs in colleges and universities. They taught the techniques of test construction they had developed in the army and adapted these to civilian educational uses. A result is that today such normed or standardized tests are widely believed to define educational testing and are often called, and believed to be, achievement tests. The following discussion shows, however, that both of these popular notions are dubious and a cause of some serious misunderstandings.

CS 201 388

We begin by explaining why standardized testing cannot be equated with educational testing generally. If you had to write a standardized test, you'd have to devise items that half the typical takers of the test could not do. If the typical student doesn't fail half the items, then the test isn't functioning as a standardized test must.

It is easy to see, then, that the basis for writing a standardized test is not what students know or can do. Rather the purpose is to determine how one student compares with another on a scale from high to low.

But schools and teaching are intended to have students know things and be able to do things. It would follow that for practical teaching, the right kind of test is one that gives the student the opportunity to display the extent to which he can do the things he has been taught to do, understand, appreciate, and so on. Such instruments of evaluation get at desired achievement. Despite this, we find that school systems typically measure and report educational growth not with tests designed to reflect specifically what students have been taught but rather with tests intended to discriminate among students.

We use the term "criterion test" or "criterion-related" or "criterion-referenced" to name tests that measure directly and specifically the intentions of teaching. In such tests, norming or standardizing is not involved. And the test discriminates only on such matters as whether the student possesses the achievements desired. To write the test is to elicit the skills taught. For a criterion test, such skills and other behaviors are the criteria of test construction. That explains such terms as "criterion-related" and "criterion-referenced," and why it is reasonable to call such instruments achievement tests.

It is easy to see, then, the dubiety of typical reports of school systems.

A school system may report, for example, that the achievement in reading of its students is good, bad or indifferent. Typically the data are based on results of this or that standardized test. For reasons given, the validity of these reports, including references to grade-level equivalents, are dubious and misleading; it is fair to conclude that such misleading reports exemplify a misuse of standardized tests.

The "Coleman report," and the "Jencks report" exemplify a second and related misuse. In the winter, 1974 issue of The Public Interest, Ralph W. Tyler notes that the Coleman report on Equality of Educational Opportunity and Christopher Jencks' book, Inequality claim that schools are relatively ineffective in teaching the disadvantaged. Tyler points out, however, that

> "Both the Coleman and the Jencks studies examined differences in scores on standard tests among different groups of children. They did not ask what different groups of children had learned but rather what measured variables [i.e. socio-economic status] were related to differences in scores. The standard tests used were norm-referenced tests. In building these tests, questions that most children could answer correctly were eliminated, but questions which only about half the children could answer correctly were retained. This was done in order to spread the scores as widely as possible so that children could be arranged on a scale from highest to lowest. The purpose of norm-referenced tests is to sort students, not to assess what they have learned. It happens that many of the items that are effective in sharply sorting students are those that are not emphasized in a majority of schools."[2]

Tyler goes on to note that by age 13, 80 per cent of American children can read and comprehend a typical newspaper paragraph. An exercise such as this is included in the National Assessment of Educational Progress. The purpose of this National Assessment is to report what proportion of children have acquired this and other useful skills that schools do teach. Such an exercise is not included in standardized tests for 13-year-old children "because it does not sharply separate the very skillful reader from others."[3] Coleman and Jencks were using these standardized tests because they show the largest differences among groups. They found that family background was more related to these differences than the effects of the school were. But neither the test data nor the method of analysis of variance that they used could answer the question of what most children had learned in school.

It follows that if we want to know how good schools and teaching are we must use tests that measure what schools and teachers teach. The tests must be so written that if the student has learned, the test will show it. Teachers can write such tests and put them to constructive uses. The principle for writing the test is simple and obvious: Begin with statements of what the student is expected to know or do. Then write items that give him the opportunity to exhibit these desired achievements. If, for example he is to be able to state the literal sense of what is going on in a lyric poem, then present him with a lyric poem that presumably he hasn't seen before but is within the range of his experience and ask him to explain what, literally, is going on in it.

Standardized tests, however, do have their uses. If there is a need for sorting students on a range from low to high, then a sorting kind of test is appropriate. College admissions tests exemplify one of these uses and the Graduate Records Examination exemplifies another.

The purpose of teaching, however, is not to sort students but rather to help them achieve. Because desired achievements are the criteria for writing criterion tests, these are proper for typical classroom uses.

To summarize: Standardized tests are intended to sort people out, not to elicit learning sought by particular teachers and schools. It is misleading, therefore, to use the results of standardized testing alone to represent the achievements of educational programs. It is also misleading to use standardized tests to compare one school or school system with another: such comparisons are not necessarily based on the skills or other kinds of behavior that any or all of the schools involved have taught. A program of instruction is properly assessed by criterion tests because these are based on the skills taught and so written that the students to be tested are given an opportunity to demonstrate the extent to which they have acquired the desired learning. The results of such tests provide the primary acceptable bases for improving our schools and our teaching.

NOTES

[1] Revised version of a talk presented to a session of the Twenty Fourth Annual Conference of the New York State English Council, Binghamton, New York, May 4, 1974.

[2] Ralph W. Tyler, "The Federal Role in Education," The Public Interest, No. 34. Winter 1974, 170.

[3] Ibid. 170-71.